DOCUMENT RESUME

ED 127 314 SP 010 351

AUTHOR Lai, Morris K.
TITLE Evaluation in a Large-Scale Curriculum Development Setting.
PUB DATE 76
NOTE 7p.; Paper presented at the meeting of the American Educational Research Association (San Francisco, California, April 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Curriculum; Curriculum Development; *Curriculum Evaluation; Curriculum Planning; Educational Development; Educational Finance; *Educational Planning; *Formative Evaluation; *Statewide Planning; *Summative Evaluation
IDENTIFIERS *Hawaii

ABSTRACT

Curriculum development in Hawaii is often implemented on a large scale, statewide basis with concentrated resources and financial support. This presents a unique setting for comprehensive evaluation. Various methods of accomplishing the evaluation and satisfying both community and legislature are discussed. A variety of evaluation resources are used. (JMF)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED127314

April 14, 1976

EVALUATION IN A LARGE-SCALE CURRICULUM DEVELOPMENT SETTING*

Morris K. Lai
University of Hawai'i

Evaluation has always been a most challenging task, and Scriven has gone so far as to characterize it as requiring almost all the skills known to man (sic.). The unique setup of the Curriculum Research and Development Group at the University of Hawai'i, however, has made evaluation both easier and more difficult. The extra difficulty is in part due to the complexity as well as to the large scale (often statewide) implementation of the programs. On the other hand, because of concentrated resources and financial support, the evaluation activities have had a sort of support that does not often exist elsewhere.

We have gained a lot of valuable experience throughout our many years of curriculum development. As time went on, our evaluation methodology underwent a steady evolution. By sharing some of the lessons we learned, it is hoped that others might benefit by either not having to re-invent the evaluation wheel or by being able to get some empirically tested methods for incorporation into their own evaluation methodology.

The attempt at comprehensive evaluation at the Curriculum Research and Development Group has been relatively recent. In the past, the State legislature has not seen fit to appropriate more than a token amount of funding for evaluation purposes. Adequate support levels for evaluation are no longer impossible, but funding is certainly not automatic, and strong cases have to be made annually in testimony before the legislature. As might be expected, the legislature and the Department of Education have been more interested in summative than in formative evaluation. Unfortunately summative evaluation is often asked for immediately even though very little formative evaluation has taken place. This state of

SP010 351

* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

affairs is hardly conducive to carrying out summative evaluations. The formative/summative distinction in our setting exhibits a pattern that is different from the distinction often made in the literature; for example, what would definitely be formative evaluation, say, in the Sanders-Cunningham (1973) schema, may be summative in that future funding depends on whether the product is effective regardless of the fact that the product has not yet been integrated into a complete program or curriculum. If one only did formative evaluation (i.e. in order to revise the product), then financial support might become shaky. On the other hand a purely summative approach would not take advantage of the potency of formative evaluation to help improve products. Attempts to accommodate such concerns have led to labels like "interim summative evaluation." This type of evaluation is similar to that performed in the interim/external evaluation cell in the Sanders-Cunningham (1973) formative evaluation model; however, the term "summative" is a crucial difference. Borich (1974) among others has indicated that the formative/summative distinction is often not clear; we have found that not only is the demarcation line often blurred, but also certain evaluation stages are both formative (to improve the product) and summative (to see if further financial support is justified). In general the curriculum developers have been somewhat uneasy about evaluation that is largely summative in nature, and it is important to emphasize that most of the evaluation efforts that concern them are geared toward a goal that developers as well as evaluators have: the improvement of the product; however, when the legislature and Department of Education ask about evaluation they almost always mean summative evaluation. Rather than simply report only summative results, it is recommended that formative results also be presented as evidence of attempts to improve the quality of the product. Often officials are pleased to receive such information, but they had simply not thought of the attempts to improve the product as being related to "evaluation."

Another specific concern is the polyglot makeup of Hawai'i, which has resulted in challenging evaluation problems; for example, how does one draw a representative sample when almost a dozen ethnic groups have substantial representation in the population? Certain test questions may be biased against one group but not another. Mainland developed tests can be (and often are) biased against Islanders. With approximately 8% of our population being immigrants, it is impossible to make valid, controlled comparisons with how the rest of the United States is doing (e.g. on standardized achievement tests), although such comparisons are constantly being made by educators, legislators, and some evaluators. Evaluators everywhere undoubtedly have somewhat similar problems. It is how the problems are attacked that may be of some interest to this audience.

Because sampling is often done at the school level, it behooves us to have relevant data that are relatively accessible on all schools. Socioeconomic (SES) data that are available include percentage of families on welfare, block statistics from the U. S. Census, occupations and salaries of heads of households, average cost of housing, etc. By having recent SES information it is possible to quickly select randomly (stratified) a group of schools that either is representative of the State or satisfies a certain composition need. Sampling at the pupil level is also facilitated by an up-to-date file of individual demographic data.

Perhaps in part because of Hawaii's isolation we have recognized a need to be current with regard to the work being done elsewhere. Literature reviews often include a retroactive computerized search by ERIC complemented by the monthly current awareness searches. Measures are located, screened, selected, or rejected with the assistance of UCLA's Center for the Study of Evaluation's comprehensive critiques of existing tests, the Robinson and Shaver (1973) compendium of tests in the social psychological area, various test summaries put out by ETS, the

National Council of Teachers of English, Research for Better Schools, Buros, etc. Objectives and item banks that are used where helpful, include those of IOX (UCLA), Westinghouse Learning Corporation, National Assessment, SOBAR, and states like California and New York. The CEDAR catalogue, as well as periodical newsletters, apprises us of the work being done at the regional laboratories and centers. The latest books on evaluation also provide us with a means of keeping up with the state of the art. Ones that have been particularly useful are Worthen and Sanders (1973), Popham (1974), Borich (1974), Walberg (1974), and Yelon (1974). And finally we attempt to get the latest in "live" advice from evaluation consultants like Scriven, Klein, Cahen, Heath, and Worthen whom we hire not only to provide a bit of externality but to also help us improve our evaluation efforts.

Now given all these attempts to make use of evaluation resources, problems still remain. Evaluation models must be eclecticized in order that they be appropriate to particular projects. Measures and objectives need to be modified or developed from scratch, bureaucratic structures must be dealt with, and the results must be explained meaningfully to various audiences.

Some success in the first area has resulted by means of our combining the best parts of the various models; for example, the Sanders and Cunningham formative evaluation model that extended Scriven's work provides a general framework that can be complemented by Hess and Wright's (1972) approach. Scriven's formative and summative checklists are then used as anchors and means of metaevaluation (also useful are the criteria suggested by Stufflebeam (1971)).

Where community or legislative involvement in the selection of objectives is important, it is often useful and more effective to present interested parties with actual items and ask them to decide whether or not they agree with the item (rather than the more general objective) as being a valid measure of success/failure or quality. Be forewarned, however, that such "sign-offs" before data

are collected do not guarantee the lack of future complaints by the various audiences (as experienced by Klein et al, 1975). Combating bureaucracy can sometimes be facilitated by liaison personnel who have connections in the necessary places. Such a person is almost like an ombudsperson and certainly can help oil the wheels for evaluations.

Finally in order to come anywhere near the level of having all the necessary evaluation skills, we rely heavily on lessons from the past and from other sources. It would, therefore, be very much appreciated if any of you would share with us some of your evaluation techniques or experiences that have proved to be valuable. Such an interchange has the potential to make evaluation for us practitioners both more pleasant and more do-able.

REFERENCES

Borich, G. D. (Ed.). Evaluating educational programs and products. Englewood Cliffs, N.J.: Educational Technology, 1974.

Buros, O. K. Mental measurements yearbooks. Highland Park, N. J.: Gryphon Press, 1938, 1940, 1949, 1953, 1959, 1965, 1972.

Hess, R. J., & Wright, W. J. Evaluation strategies as a function of product development stages. Paper presented at the American Educational Research Association, Chicago, 1972.

Klein, S., Hoepfner, R., & Kosecoff, J. A progress evaluation of a national ETV series. Paper presented at the American Educational Research Association, Washington, D. C., 1975.

Popham, W. J. (Ed.). Evaluation in education. Berkeley: McCutchan, 1974.

Robinson, J. P., & Shaver, P. R. Measures of social psychological attitudes. Ann Arbor: Institute for Social Research, University of Michigan, 1973.

Sanders, J. R., & Cunningham, D. J. A structure for formative evaluation in product development. Review of Educational Research, 1973, 43 (2), 217-236.

Walberg, H. J. (Ed.). Evaluating educational performance. Berkeley: McCutchan, 1974.

Worthen, B. R., & Sanders, J. R. Educational evaluation: Theory and practice. Worthington, Ohio: Charles A. Jones, 1973.

Yelon, S. L. Constructive evaluation -- Improving large scale instructional projects. Lansing: S. L. Yelon, 1974.